AD_____

GRANT NUMBER DAMD17-96-1-6229

TITLE: Cancers Missed on Mammography

PRINCIPAL INVESTIGATOR: Robert Schmidt, M.D.

CONTRACTING ORGANIZATION: University of Chicago
Chicago, IL 60637

REPORT DATE: October 1997

TYPE OF REPORT: Annual

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;
distribution unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| 1. AGENCY USE ONLY *(Leave blank)* | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | October 1997 | Annual (23 Sep 96 - 22 Sep 97) |

**4. TITLE AND SUBTITLE**
Cancers Missed on Mammography

**5. FUNDING NUMBERS**
DAMD17-96-1-6229

**6. AUTHOR(S)**
Robert Schmidt, M.D.

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
University of Chicago
Chicago, IL 60637

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

**11. SUPPLEMENTARY NOTES**

**12a. DISTRIBUTION / AVAILABILITY STATEMENT**
Approved for public release; distribution unlimited

**12b. DISTRIBUTION CODE**

**13. ABSTRACT** *(Maximum 200*

Our goal is to decrease the estimated 30% of breast cancers missed on screening mammograms by radiologists, by applying computer-aided diagnosis (CAD) techniques to mammogram interpretation. A database of 100 screening mammograms containing cancers that were observational misses in routine practice will be collected and analyzed for pathology, mammographic lesion type, breast density, location, subtlety, size, growth rate, and reasons for having been overlooked. These cases will be digitized along with 300 normals, and presented to 12 general radiologists and 3 experts, to evaluate observer performance, once with and once without CAD prompts pointing at suspicious areas. The radiologists' readings will be compared for sensitivity, specificity, and positive and negative predictive value. The hypothesis is that CAD can decrease the number of misses by up to 50%, and that we can measure this using a unique database of cancers containing a large fraction of cancers that a radiologist might miss without CAD help. In evaluating interobserver variability of readers, we anticipate that CAD second opinions can help average readers perform closer to experts. The results will be used to justify and plan the future implementation of CAD in the form of commercial products for practical use in clinical mammographic practice.

| 14. SUBJECT TERMS | 15. NUMBER OF PAGES |
|---|---|
| Breast Cancer; Mammography | 25 |
| | 16. PRICE CODE |

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | Unlimited |

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

RM Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

RM Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

RM For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

11/30/97

PI - Signature Date

# TABLE OF CONTENTS

## 1.1 Subject and scope of the research work

Screening mammography has been conclusively shown in randomized prospective trials to be the most effective method to reduce the enormous toll of breast cancer mortality, by 30% or more. Mammography is the only known screening method available that can detect breast cancers at an early enough stage to significantly affect the patient's outcome (ref. 1). Much effort is currently being devoted to developing alternative screening methods and improving the imaging technique of mammography, such as making the acquisition and display process digital. In the current screening efforts, and probably in most future imaging efforts relying on human observers, the radiologist reading the mammogram is probably the most variable component, and potentially the "weak link" in screening programs. Early mammographic signs of breast cancer such as small masses, distortions and clustered microcalcifications are often very subtle changes, and missed even by highly trained radiologists 10 to 30% or more of the time in the published literature. The reasons for such misses vary, but have been categorized as due to low conspicuity, eye fatigue and simple human oversight (refs. 2-4). It is the fundamental assumption of this research that improvements tin early breast cancer detection hat are very significant and comparable in magnitude to most improvements that can be achieved through the development of new imaging modalities can be achieved by improving the effective abilities of the observer. Large sums are being spent to support research into alternate modalities for detecting breast cancer, and these more costly detection methods might not add much compared to improved mammographic performance. In particular, this research focuses on the improvements to be obtained by providing to the radiologist interpreting the mammographic images the assistance of computer prompts for potentially significant abnormalities, or Computer-aided Diagnosis (CAD).

Double reading of mammograms has been shown to significantly increase the number of cancers detected (ref. 5), and it has been proposed that a computer might act as a second reader, using a form of CAD (ref. 6). For CAD to be effective (1) computers must find cancers that are missed by radiologists, and (2) radiologists must react appropriately to the computer prompts. We have found that computer detection schemes can find 50% of the observational misses made by radiologists reading mammograms (ref. 7). The answer to the question of whether radiologists will correctly use the prompting information generated by CAD is unknown, although several studies using cases found by radiologists show promise for improving observer performance for spiculated masses and calcifications (refs. 8,9).

The current study is designed use a large database of cancers already missed by radiologists in routine clinical practice, and will test observers without and with the aid of CAD. It is expected that radiologists will detect about 10 to 15% more cancers using CAD, which would have important implications for bringing this technique into clinical practice. We will also learn much more about the reasons for and types of radiologist misses on mammography. The size of the database (400 cases, with 25% being cancers already missed in routine clinical practice) and the number of observers (12 general radiologists reading mammograms) distinguish the current work from prior studies, as well as the inclusion of multiple types of lesions representing the range of what is seen in normal screening practices. The ideal situation would be to do a prospective clinical study rather than a retrospective observer study, but the low incidence of breast cancer in typical screened populations (about 5/1000), and the difficulty of obtaining truth, including long term follow up to determine the false negative rate, makes this impractical. Using a database constructed to be difficult by using known cancers already missed in clinical practice represents a reasonable solution to these problems, making such a study of CAD's effect feasible.

Another purpose of this study is to obtain much needed data on the inter-observer variability encountered in everyday practice, as some studies have indicated this effect may be quite large (ref. 10). Lastly, this can add to the understanding of the types and reasons for radiologist misses on mammography, which have varied in previous investigations (refs. 11,12).

The University of Chicago Kurt Rossmann Laboratories for Radiologic Image Research is a group of researchers headed by Dr. Kunio Doi, PhD, who have worked for over a decade developing the concepts and practical methods of CAD with which to assist radiologists in detecting and analyzing lesions seen on various types of imaging studies. A major concentration has been on mammography and improving the sensitivity of breast cancer diagnosis (ref. 13).

## 1.2 Purpose

The goal of this project is to demonstrate the clinical usefulness of CAD in mammographic screening, by showing that radiologists can detect more breast cancers and/or earlier breast cancers than occurs now. This is to accomplished by collecting and refining a dataset of lesions that radiologists are known to have missed in clinical practice, and applying CAD methods to the digitized film data The presentation of CAD prompts to radiologists reading the mammograms should result in a performance increase comparable to studies involving

second human readers, and to prior smaller studies utilizing selected databases of specific lesion types. The performance increase expected is therefore about 15% or more, on a database that is known to include cases which have already been shown to give radiologists difficulty. The major hypothesis is that significantly more breast cancers will be detected by radiologists who have access to CAD "second opinion" outputs for mass lesions and calcifications, by reducing observational errors in reading the mammographic images.

The specific aims of this project are:

(1) Establishing a database of 100 cases of observational misses of breast carcinoma, and categorizing the lesions as to type and reason for miss in a uniform fashion. Another 300 cases of normals selected to represent the range of what is seen in clinical practice will be collected to establish the test dataset to be presented to observers.

(2) Digitizing all the above mammogram cases and running CAD algorithms for masses and calcifications developed at the University of Chicago, to provide graphic output of the potential location of lesions in both the normal and abnormal cases.

(3) Presenting the 400 cases to 12 observers both without and with CAD output, and recording their sensitivity and specificity of cancer diagnosis, using ROC methods for analysis. The method of presentation will be developed and refined over the course of the project.

(4) Analysis of the effect of CAD in terms of observer performance benefits in a simulated clinical situation, using ROC analysis.

## 1.3 Background of previous work

Sensitivity in detecting early breast cancer is mammography's most important feature. While mass screening mammography has been shown to significantly decrease mortality from breast cancer, the large scale studies that have been done primarily compare results to those in control populations, and do not state the sensitivity of the test except relative to physical examination. For example, in the Breast Cancer Detection Demonstration Project (BCDDP), mammography was shown to significantly outperform its closest competitor, physical examination. However, if one were to include as potentially detectable all interval cancers

diagnosed within one year of a screening, the sensitivity of detection for mammography in the BCDDP is 80% (ref. 14). As experience with screening mammography has grown in the US, some of these limitations of mammography have become more evident. Potentially the most serious of these is the fact that mammographic interpretations do not report all the potentially detectable cancers, due to limitations of the observer. It is not known whether the average radiologist, who does not specialize in mammography, can reproduce in routine clinical practice the literature-reported successes of mammography. It can be estimated that 30% or more of potentially detectable lesions are missed, with miss rates even by experts of 10 to 15%. This is corroborated by reports from authors who have audited their work, in which experienced readers have sensitivities of 85-91% (refs. 12,15,16). Given the growing acceptance of mammographic screening in the US by health care providers, the public at large, and managed care organizations, coupled with the increasing population for which mammography is appropriate, the importance of reducing the error rate is increasing.

Very little is known about the sensitivity of average radiologist observers in routine clinical practice. In fact, most mammograms performed in this country are interpreted by such average observers, the majority of whom had relatively little mammography training during their residency years. It is equally true that mammography has begun to feel a backlash due to its imperfections, even before it has achieved universal acceptance. Part of this is due to the fact that mammography training, and in particular training in the screening function, is a relatively recent phenomenon. The detection of spiculated lesions, camouflaged in the nodular background of varying densities in the parenchymal tissue structures, is a task that requires experience, attention to detail and perceptual acumen. Estimates of the number of missed breast cancers vary, but range from about 10 to 30 % (17-20), with estimates as high as 70% on retrospective rather than prospective review (11).

The screening task is one of *detection* of potential abnormalities, with diagnostic mammography used for *analysis* of the malignant potential of a lesion, once it has been located. The logistics of training for detection of abnormalities in a way that simulates the clinical experience are formidable, when over 99% of screening mammograms do not show a cancer. The repetitive task of looking for subtle abnormalities on screening mammograms can be likened to assembly line work in industry, which lends itself well to automation. Radiologists do not identify all breast cancers visible on mammograms, and despite their extensive training, do not necessarily perform better than physicians' assistants who have undergone a short course of intensive training in the basic search patterns used to detect

suspicious abnormalities in mammograms (ref. 21). The detection task has to be efficient, distinguishing the myriad of normal structures from those infrequent breast tissue patterns which signal a malignancy. Moreover, average readers interpreting mammograms as part of their ordinary clinical practice may be much less sensitive than the reports from experienced or expert readers indicate. Although double reading has been shown to improve detection rates, this practice is relatively unusual in this country, and logistically is unlikely to become the standard using a second human observer. Recent developments in CAD using digitized mammograms have indicated that about half of the observer misses on mammography can be successfully flagged by currently available computer programs (ref. 7), potentially reducing the miss rate by a significant proportion. Computers are consistent, tireless, remember the past perfectly, and do not complain when asked to work long hours. They do not suffer from the human foibles of irritation and distractibility, and they do not spend time worrying about their decisions. With the help of a computer that points at and, hopefully, ultimately aids in analyzing lesions, average radiologists could be expected to perform at levels near that of experts.

In order to improve detection rates in mammography and bring down the false-positive and false-negative error rates, a great deal of work is being done using computers (refs. 6-9,22-31). Over the past decade at The University of Chicago, we have been developing CAD programs to detect potential masses and calcifications on mammograms, with the goal that this system may eventually act as a second reader. The aim of research into the use of such methods is not to see if digital (or digitized) mammography and computers can perform better than humans at reading mammograms, but instead to determine whether they can enhance human performance and reduce the number of mammographic misses. This proposal is designed to study the inter- and intra-observer variability of radiologists in practice and to determine if their performance on a unique database of cancers already missed in routine clinical settings can be improved by the addition of CAD. The database constructed will bring together missed cancers and a larger number of normal mammograms from The University of Chicago files, and from a well-audited, large private practice in New Mexico (refs. 16,32). A preliminary study has been done that shows the potential for implementing CAD in routine clinical practice (ref. 7) and the relative success of computer detection of missed cancers, even when they are relatively subtle (ref. 28). The purpose of the proposed study would be to quantify the performance of radiologists in reading with and without the benefit of CAD. The objective is to improve the accuracy of schemes that provide computerized detection signals and evaluate critically their effectiveness as an actual decision aid for the mammographer, in complex screening situations with subtle lesions. The evaluation phase is essential to test how effective current computer

vision techniques are, and to determine their optimum use and magnitude of benefit in clinical practice. Detailed evaluations, such as the one proposed here, have not be carried out on large databases such as the one we will assemble. Based on the results of this study, this could lead to justification for implementation on a wide scale of CAD techniques in screening mammography, and in particular, provide a quantification of the effective benefit to the human observer. Ultimately, this could lead to a method of decreasing the miss rate in mammography by 25% or more, with no further testing of the individual patient and at modest capital investment.

The causes of missed lesions on mammography are varied. Observer errors of (1) observation, (2) interpretation and/or (3) communication constitute categories of missed lesions that can be minimized by knowledge and experience, combined with consistent application of an effective, systematic approach to mammographic image-reading tasks. In our experience, observation errors constitute the single largest category of missed lesions. By studying eye-positions and search strategies of radiologists looking for lung nodules on chest x-rays, Kundel, et al. (ref. 33) identified 3 possible causes of diagnostic errors: (1) inadequate search (30%), (2) failure to detect/recognize (25%), and (3) faulty interpretation/decision (45%). The first two of these error types (total 55 %) are observational type misses. It is reasonable to assume that misses in mammographic searches are similar. Analyzing cancers missed on screening mammography, Bird, et al. (ref. 12) found that 43% of misses were due to the lesions being overlooked, 52% were due to misinterpretation, and 5% were due to suboptimal technique. Individual practitioners' skills (not published accounts of other radiologists' abilities) are what make screening mammography a success or failure. There is much to learn by looking back in time on previous mammograms to see if there are any perceptible signs of the developing abnormality on the prior studies. Such signs probably are present in over 1/3 of the cases of cancers detected on mammograms. Retrospective reviews enable the radiologist to see the subtlest signs of the forming cancer in many cases, and this method of review is of course more sensitive than blinded prospective review of the same cases (ref.11). One can thereby increase understanding of the manifestations of the earliest forms of breast cancer, and achieve an ability to recognize potential cancers before they reach their more obvious stages.

CAD can be defined as a diagnosis made by a radiologist who takes into consideration the output of an automated image analysis when making his/her decision. The University of Chicago has pioneered the application of computer methods to improve mammogram interpretation. Over the past ten years, CAD programs to detect potentially suspicious areas of

calcifications and masses on digitized mammograms have been developed and verified by observer testing, achieving detection rates of 85 to 90%. Observer performance studies have documented improvement in radiologist performance (Table 1). At the current state of development, the computer can be asked to point out potential lesions on digitized conventional screen film mammograms, using programs developed to detect masses and calcifications. We have applied this method to lesions which were observational misses by radiologists. The computer was able to detect half of these, and the ability to detect cancers does not seem to be strongly dependent on the subtlety of the lesion (refs. 7,28). In effect, CAD acts as a friendly "second reader." It is still up to the radiologist to decide if the suggested areas merit further evaluation. We installed the first clinical CAD system in our department in late 1994 (29). The hope is that it can be perfected to serve as a routine clinical tool and ultimately enhance our detection accuracy on screening mammograms.

Lastly, radiologists are increasingly being accused of malpractice in missing breast cancer, which has emerged as the leading cause of suits (Physicians Insurers Association of America , Breast Cancer Study, June 1995, Washington, DC). This can have a deleterious effect on screening mammography, by artificially increasing patient call back rates for additional work up by radiologists due to fear of missing cancer. In the long run, radiologists may be able to rely on the CAD results to support their decision that a mammogram is normal, but this would only be after CAD has undergone extensive clinical testing and been shown to be reliable. The real problem is that the miss rate on screening mammography is too high, particularly among average radiologists, but there is not much data to quantify this. Initial missed lesion analysis is evidence that supports CAD as a solution, but we are not sure. Clear evidence that radiologists' accuracy can be improved when the computer points out lesions for them has been achieved in the laboratory setting, but it is not known to what extent this can be realized in the context of clinical practice, where the number of normal mammograms far exceeds those showing a cancer (ratio 200:1). Obtaining evidence that state of the art CAD programs can achieve clinical utility is the crucial next step in deciding on the proper implementation and future development of this technique. The University of Chicago has had much success with computerized detection schemes for mammography. For microcalcification clusters we have achieved a true-positive (TP) rate of 85% with a false-positive (FP) rate of 0.5 per image (ref. 26). For masses (ref. 25), our detection scheme yields a sensitivity of 92% with an average of 2 FPs per image . Note, however, that if the computerized detection program were to be used in its current state as a screening tool in the clinic, almost every image would be "flagged" as potentially containing a lesion because of the program's false-positive rate. Thus, in order to make a computerized

detection program more clinically effective for the mammographer, its overall detection accuracy must be assessed. We have found that in our preliminary trials of over 2000 screening cases, the radiologist is able to dismiss relatively easily the false positives generated by the computer. It remains to be shown whether leading the radiologist horse to the water will result in drinking; it is possible that radiologists may not sufficiently react to the computer prompts if the patterns of disease are not familiar to them and recognized. Available data based on improvements in ROC performance have shown that CAD will improve radiologists' performance on lesions that other radiologists have detected. The proposed study is designed to answer a crucial question: whether CAD can be equally effective in enhancing radiologists' performance in the case of lesions which radiologists in their routine practices have failed to detect. Even relatively modest improvements in performance in this domain would translate into realization of clinical benefits, as the delay times are often significant, causing tumors to be detected at stages where they are less curable. The average doubling time of tumors missed in our preliminary study was about 300 days, with a range from 30 days to 1500 days.

# 2. BODY OF REPORT

## 2.1 Experimental methods, assumptions and procedures

The project will be accomplished using a database of accumulated clinical observational errors in breast diagnosis, by presenting films to observers with and without the benefit of CAD output. We propose to construct a mammogram database of 100 missed cancers and 300 normals and test 12 experienced but not expert radiologists, presenting the cases over an extended period both with and without access to CAD output. Each radiologist will review the database of 100 missed cancers and 300 normal cases twice, once with and once without CAD output, at times sufficiently separated to negate any memory effects. The inter- and intra-observer performance of these radiologists will be obtained from the testing, and compared with the limited previously published data, and additional observer data accumulated as part of the first year of this project (see 2.2.2 below). This concentration of cases that are difficult for radiologists should provide the necessary foundation for deciding on how quickly and aggressively to pursue this modality. Additionally, the promise of direct digital mammography is on the horizon. Computer analysis of images produced by that technique would be a relatively inexpensive adjunct that should be planned for, given the encouraging preclinical CAD results, and the relative height of the ceiling for improvements that are now being investigated. With the increasingly powerful computers and sophistication in artificial intelligence methods, this type of development to aid diagnostic decision making is inevitable, in our opinion, and has reached the point where it needs to be critically evaluated clinically. A secondary benefit of the proposed investigation will be valuable information on which characteristics of cancers make them more difficult to detect, and for which types CAD provides the biggest gain. Verification of an expected significant improvement in breast imaging interpretation and elevation of the performance of average observers to a level approaching that of experts is the goal.

The hypothesis of this project is that computer-aided diagnosis (CAD) can reduce the observation miss rate of radiologists reading mammograms by up to 50%. This study is based on the assumption that the specific tumors that radiologists miss in their daily work will provide a unique and rich database to test the real clinical potential of CAD at its current level of development, and will provide a basis for enhancements in the future. It will also provide a clear-cut range of detection rates for lesions that are relatively subtle (i.e., those which tend to be difficult for human observers) and can therefore be used to measure inter-observer variability. It may also provide insight into the structuring of standard tests for observer performance in mammography, an issue which is beginning to receive attention but for which

no accepted standard has been developed. Further study of the differences between the human and the computer observer may be fruitful, in that analysis of the differences in errors made by one versus the other can be analyzed. It is also expected that the characteristics of the tumors missed frequently by radiologists in such a setting will be useful both for structuring training programs for radiologists and for improving computer detection schemes in the future.

The prime technical objective of this project is to obtain a relative sensitivity for detection of breast cancers missed in clinical practice for radiologists reading with the benefit of CAD output, compared to the same radiologists reading without the benefit of CAD.

A secondary technical objective is to evaluate inter- and intra-observer performance variations, with and without CAD.

The third technical objective is to accumulate clinical characteristics and significant features of the 100 missed cancers that will be entered into the missed lesion database.

The fourth technical objective is to evaluate radiologists reactions to the use of CAD in a simulated clinical environment, including preferences for display modality of the data.

## 2.2 Results and Discussion

Over this initial period of the grant, three main goals have been achieved:

### 2.2.1 Development of database of missed lesions for observer study

The collection and categorization of the missed lesions essential for the study has been about two-thirds accomplished, with about 50 to 60 missed malignant lesions from each of the two sites (Univ. of Chicago, X-ray imaging Associates of New Mexico) identified. Those from the University of Chicago have been categorized and digitized, while those from Dr. Michael Linver's practice in New Mexico are scheduled in several sessions in the next 2 months to be fully categorized after preliminary categorization, and digitized by the research assistant for the project. CAD results have been run for the University of Chicago cases.

### 2.2.2 Preliminary observer study in a simulated screening environment

A major effort was placed this past year on accumulating data on "average" radiologists, to help in case selection, baseline performance data, and observer study design. Preliminary results have been analyzed on the first 100 of 140 observers, who viewed 100 cases, about half of

which were cancers, in a simulated screening environment at 3 mammography courses run by the University of Chicago, New York University and Dr. Michael Linver's practice. Cases were accumulated from the University of Chicago and Dr. Linver's practice, with some cases contributed by Dr. Gillian Newstead of NYU. Four experts were also tested in this environment. The average sensitivity of general radiologists reading mammograms was about 70% for the cancers presented, with the experts average sensitivity being about 85%. To date 30,000 data points have been entered into a spreadsheet by a medical student working on the project, detailing the results. A sample of the case difficulty rankings for the cancers is given in Table 1, at one of the sessions. This information will be used to help in selecting cases for the final observer study, as cases were clearly identified that more than half the radiologists missed. Additionally, second lesions and axillary tail lesions showed excess propensity to being overlooked. It is being considered that some of these cases be included in the final database, as we have excellent categorization of cases in this series which are frequently missed. Design of the observer form has undergone several revisions, with the latest version displayed in Figure 1. The last session of 40 observers was the first to employ ROC type grading with a scale of 1 to 10, and it is currently being analyzed. At this time, it would appear that both traditional sensitivity and specificity, with lesion location designated by the observer, as well as ROC type analysis of responses, will be useful in this project. Lastly, results showing a real but relatively weak correlation with observers' self assessments of their degree of experience were accumulated, which have provided insights into how to accumulate better indirect categorization of observers' degree of expertise (Figure 2). Interestingly, one of the observers who categorized himself as advanced scored among the lowest in sensitivity in this exercise, less than 40%

### 2.2.3 CAD (Computer-aided Diagnosis) results on preliminary missed lesion database

Using the latest computer algorithms developed at the University of Chicago, Dr. Nishikawa has run the detection schemes for both clustered microcalcifications and masses on a set of 100 missed cases, which will be used as the basis for finalizing the codes to be used for the missed lesion database for this project. Analysis of this phase is in progress at this time.

## 2.3 Recommendations in relation to the Statement of Work

The overall plan of this three-year project involves three major steps: (1) assembly, preparation, digitization and cataloging of the database to be used in the second part; (2) observer performance studies testing 12 non-expert and 3 expert radiologists, presenting the cases with

and without CAD; and (3) data analysis of the results of testing. The first phase is approximately on schedule, having been estimated to take approximately a year. Elements of the second phase have been trialed in a novel way by accumulating data on a test set for a very large number of observers which is providing very valuable results to design and successfully complete the observer testing of the final missed lesion database. preliminary data analysis of these results has been accomplished, although originally it was not intended to test observers until later in the project.

As progress follows the statement of work and expectation is that the project will be completed within the originally specified grant period, no specific changes in the original statement of work are proposed.

## 3. CONCLUSIONS

The work to date follows the original proposal, with addition of considerable observer data that was accumulated on a separate database that will substantially improve the quality of the final observer study when conducted. Continued progress on development of CAD schemes is ongoing in the Kurt Rossmann Laboratories at the University of Chicago, and this also will benefit the final implementation of the algorithms employed to flag missed lesions for the observers in the experimental study which will begin next year. It is expected that these data will ultimately provide a sound basis for the introduction of CAD methods into screening mammography programs, to reduce the number of observational misses by radiologists.

# 4. REFERENCES

1. Feig SA: Decreased breast cancer mortality through mammographic screening: results of clinical trials. Radiology 167:659-665, 1988

2. Martin JE, Moskowitz M, Milbrath JR: Breast cancers missed by mammography. AJR 1979;132:737

3. Holland R, Mrvunac M, Hendricks JH, Bekker BV: So-called interval cancers of the breast. Pathologic and radiographic analysis. Cancer 1982;49:2527-2533

4. Andersson I: What can we learn from interval carcinomas? Recent results in Cancer Research 1984;90:161-163

5. Thurfjell EL, Lernevall KA, Taube AAS: Benefit of independent double reading in a population-based mammography screening program. Radiology 1994; 191:241-244

6. Giger ML: Computer-aided diagnosis. In Haus AG, Yaffe MJ (eds): Syllabus: A Categorical Course in Physics. Technical Aspects of Breast Imaging, pp 287-302. Oak Brook, IL, RSNA Publications, 1994

7. Schmidt RA, Nishikawa RM, Schreibman K, et al.: Computer detection of lesions missed by mammography, in Gale AG, Astley SM, Dance DR, Cairns AY (eds), Digital Mammography, Proceedings of the 2nd International Workshop on Digital Mammography, York, England, 10-12 July 1994 (Elsevier, Amsterdam, 1994), pp. 289-294

8. Kegelmeyer WP, Pruneda JM, Bourland PD, Hillis A, Riggs MW, Nipper ML: Computer-aided mammographic screening for spiculated lesions. Radiology 1994; 191:331-337

9. Chan HP, Doi K, Vyborny CJ, et al. : Improvement in radiologists' detection of clustered microcalcifications on mammograms: The potential of computer-aided diagnosis. Invest Radiol 1990; 25:1102-1110

10 Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretation of mammograms. NEJM 1994;331:1493-1499

11. Harvey JA, Fajardo LL, Innis CA. Previous mammograms in patients with impalpable breast carcinoma: retrospective vs blinded interpretation. AJR 1993;161:1167-1172

12. Bird RE, Wallace TW, Yankaskas BC: Analysis of cancers missed at screening mammography. Radiology 1992;184:613-617

13. Doi K, Giger ML, Nishikawa RM, et al.: Digital radiography: a useful clinical tool for computer-aided diagnosis by quantitative analysis of radiographic images. Acta Radologica 1993;34:426-439

14. Schmidt RA, Metz CE: Sensitivity of mammography (letter). AJR 1990;154:419-420.

15. Sickles EA, Ominsky SH, Sollitto RA, Galvin HB, Monticciolo DL. Medical audit of a rapid-throughput mammography screening practice: methodology and results of 27,114 examinations. Radiology 1990;175:323-327

16. Linver MN, Paster SB, Rosenberg RD, Key CR, Stidley CA, King WV. Improvement in mammography interpretation skills in a community radiology practice after dedicated teaching courses: 2-year medical audit of 38,633 cases. Radiology 1992;184:39-43

17. Houn F, Brown ML: Current practice of screening mammography in the United States: data from the national survey of mammography facilities. Radiology 1994;190:209-215

18. Baines CJ, Miller AB, Wall C, McFarlane DV, Simor IS, et al. Sensitivity and specificity of first screen mammography in the Canadian national breast screening study. Radiology 1986;160:295-298.

19. Baines CJ, McFarlane DV, Miller AB. The role of the reference radiologist: estimates of interobserver agreement and potential delay in cancer detection in the national cancer screening study. Invest Radiol 1990;25:971-976

20. Peeters PH, Verbeek AL, Hendriks JH, Holland R, Mravunac M, Vooijis GP. The occurrence of interval cancers in the Nijmegen screening programme. British Journal of Cancer 1989;59:929-932

21. Hillman BJ, Fajardo LL, Hunter TB, et al. Mammogram interpretation by physician assistants. AJR 1987;149:907-911

22. Wu Y, Giger ML, Doi K, Vyborny CJ, Schmidt RA, Metz CE: Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. Radiology 1993; 187:81-87

23. Jiang Y, Nishikawa RM, Wolverton DE, et al.: Computerized feature analysis for classification of malignant and benign clustered microcalcifications at mammography. Accepted for publication in Radiology, 1995

24. Giger ML, Vyborny CJ, Schmidt RA: Computerized characterization of mammographic masses: analysis of spiculation. Cancer Lett 1994; 77:201-211

25. Giger ML, Lu P, Huo Z, et al.: CAD in digital mammography: computerized detection and classification of masses, in Gale AG, Astley SM, Dance DR, Cairns AY (eds), Digital Mammography, Proceedings of the 2nd International Workshop on Digital Mammography, York, England, 10-12 July 1994 (Elsevier, Amsterdam, 1994), pp. 281-287

26. Nishikawa RM, Jiang Y, Giger ML, et al.: Performance of automated CAD schemes for the detection and classification of clustered microcalcifications, in Gale AG, Astley SM, Dance DR, Cairns AY (eds), Digital Mammography, Proceedings of the 2nd International Workshop on Digital Mammography, York, England, 10-12 July 1994 (Elsevier, Amsterdam, 1994), pp. 13-20

27. Bick U, Giger ML, Schmidt RA, Nishikawa RM, Wolverton DE, Doi K: Computer-aided diagnosis in mammography: Results in 100 consecutive patients with non-palpable breast cancer. Europ Radiol (Suppl) 1995; 5:S5 (abstract)

28. Schmidt RA, Nishikawa RM, Osnis RB, Schreibman KL, Giger ML, Doi K: Computerized detection of lesions missed by mammography. in 3rd International Workshop on Digital Mammography (eds K Doi, ML Giger, RM Nishikawa, RA Schmidt). Amsterdam, Elsevier Science (Excepta Medica International Congress series), pgs. 105-110, 1996

29. Nishikawa RM, Haldemann RC, Papaioannou J, et al.: Initial experience with a prototype clinical intelligent mammography workstation for computer-aided diagnosis. Proc SPIE 1995;2434:65-71

30. Schmidt RA, Wolverton, DE, Vyborny CJ: Computer-aided diagnosis (CAD) in mammography. In Kopans DB, Mendelsohn EB (eds): Syllabus: A Categorical Course in Breast Imaging (in press, RSNA publications, Oak Brook, IL, 1995)

31. Comstock CE, Giger ML, Nishikawa RM, et al. Two-year clinical evaluation of computer aided diagnosis (CAD) in the detection of breast cancer. Radiol 1997;205(P):275 (abstract)

32. Linver MN, Osuch JR, Brenner RJ, Smith RA: The mammography audit: a primer for the mammography quality standards act (MQSA). AJR 1995;165:19-25

33. Kundel HL, Nodine CF, Carmody DP. Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. Invest Radiol 1978;13:175-181.

# 5. APPENDICES

POSITIVE CASES SORTED BY NUMBER OF OBSERVERS WHO ANSWERED CORRECTLY

| CASE# | 1.1 | 1.2 | 1.3 | 2 | 3 | 4 | (blank) |
|---|---|---|---|---|---|---|---|
| 86 | 83 | 14 | 1 | | | 1 | 1 |
| 57 | 81 | 10 | 6 | | | 3 | |
| 62 | 80 | 15 | 3 | | | 1 | 1 |
| 80 | 80 | 15 | 2 | | | 3 | |
| 11.2 | 78 | 14 | | | | 7 | 1 |
| 8 | 77 | 15 | 1 | | | 7 | |
| 44 | 77 | 18 | | | | 5 | |
| 88 | 77 | 16 | 1 | | | 5 | 1 |
| 18 | 76 | 14 | 1 | | | 9 | |
| 43.1 | 76 | 14 | | | | 10 | |
| 96 | 76 | 14 | 4 | | | 4 | 2 |
| 23 | 74 | 17 | 3 | | | 4 | 2 |
| 63 | 74 | 17 | | | | 8 | 1 |
| 41.2 | 70 | 1 | 2 | | | 27 | |
| 21 | 68 | 20 | 6 | | | 5 | 1 |
| 53 | 68 | 20 | 4 | | | 8 | |
| 3 | 67 | 10 | 7 | | | 16 | |
| 31 | 67 | 12 | 9 | | | 12 | |
| 6 | 65 | 13 | 9 | | | 13 | |
| 36 | 65 | 13 | 1 | | | 21 | |
| 83 | 65 | 16 | 6 | | | 13 | |
| 93 | 65 | 16 | 3 | | | 14 | 2 |
| 19 | 64 | 11 | 10 | | | 15 | |
| 24 | 64 | 15 | 7 | | | 12 | 2 |
| 64 | 64 | 20 | | | | 14 | 2 |
| 99 | 63 | 17 | 3 | | | 14 | 3 |
| 27 | 61 | 16 | 14 | | | 9 | |
| 41.1 | 61 | 14 | 1 | | | 24 | |
| 98 | 61 | 18 | 9 | | | 8 | 4 |
| 4 | 60 | 12 | 7 | | | 20 | 1 |
| 74 | 60 | 8 | 7 | | | 25 | |
| 81 | 60 | 12 | | | | 28 | |
| 87 | 60 | 12 | 2 | | | 25 | 1 |
| 30 | 59 | 16 | 8 | | | 16 | 1 |

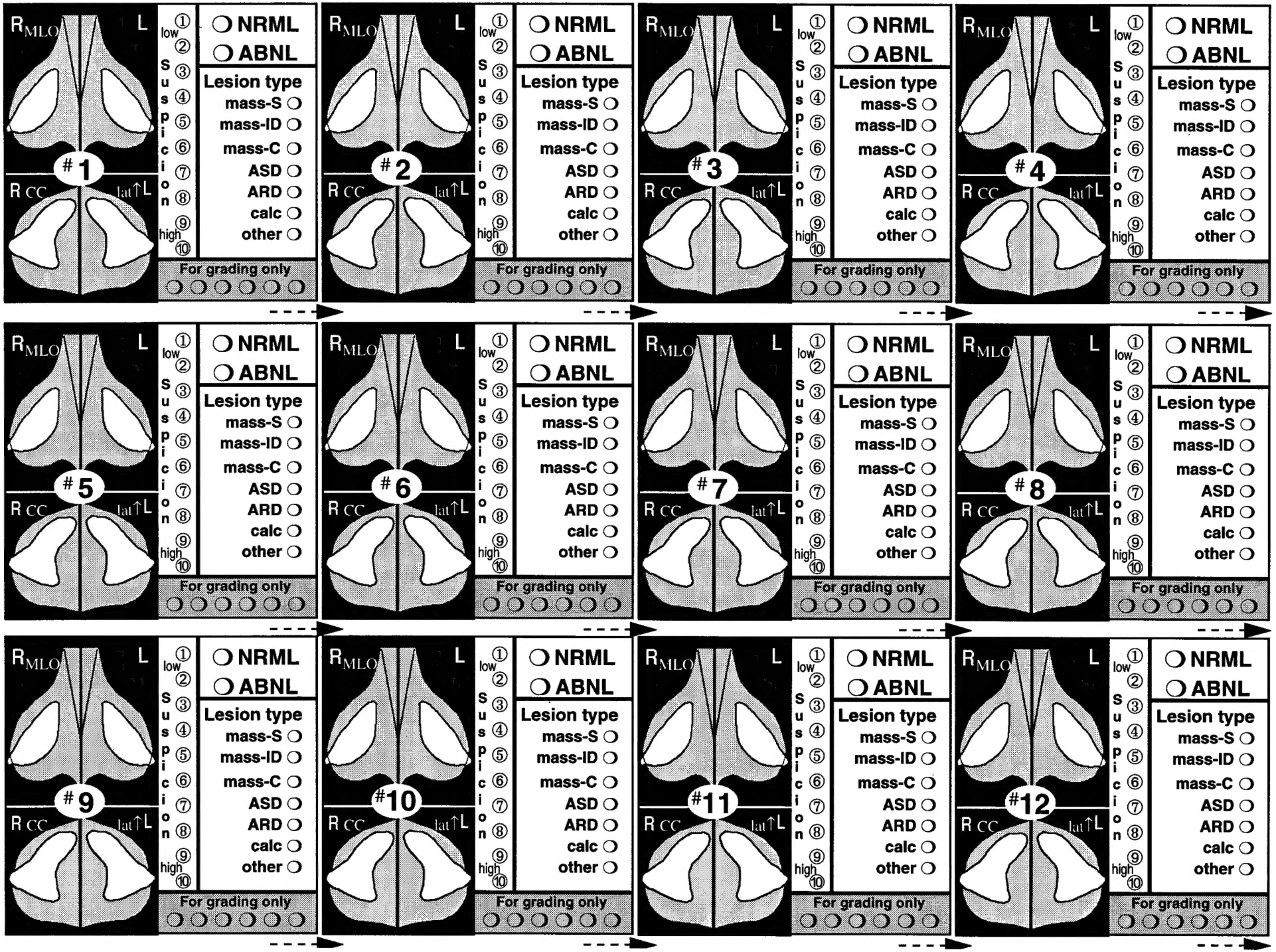
| code number | meaning |
|---|---|
| 1.1 | TRUE POS, CORRECTLY IDENTIFIED LESION |
| 1.2 | TRUE POS, NO POSITION INDICATED |
| 1.3 | TRUE POS, WRONG LOCATION INDICATED |
| 2 | TRUE NEGATIVE |
| 3 | FALSE POSITIVE |
| 4 | FALSE NEGATIVE |

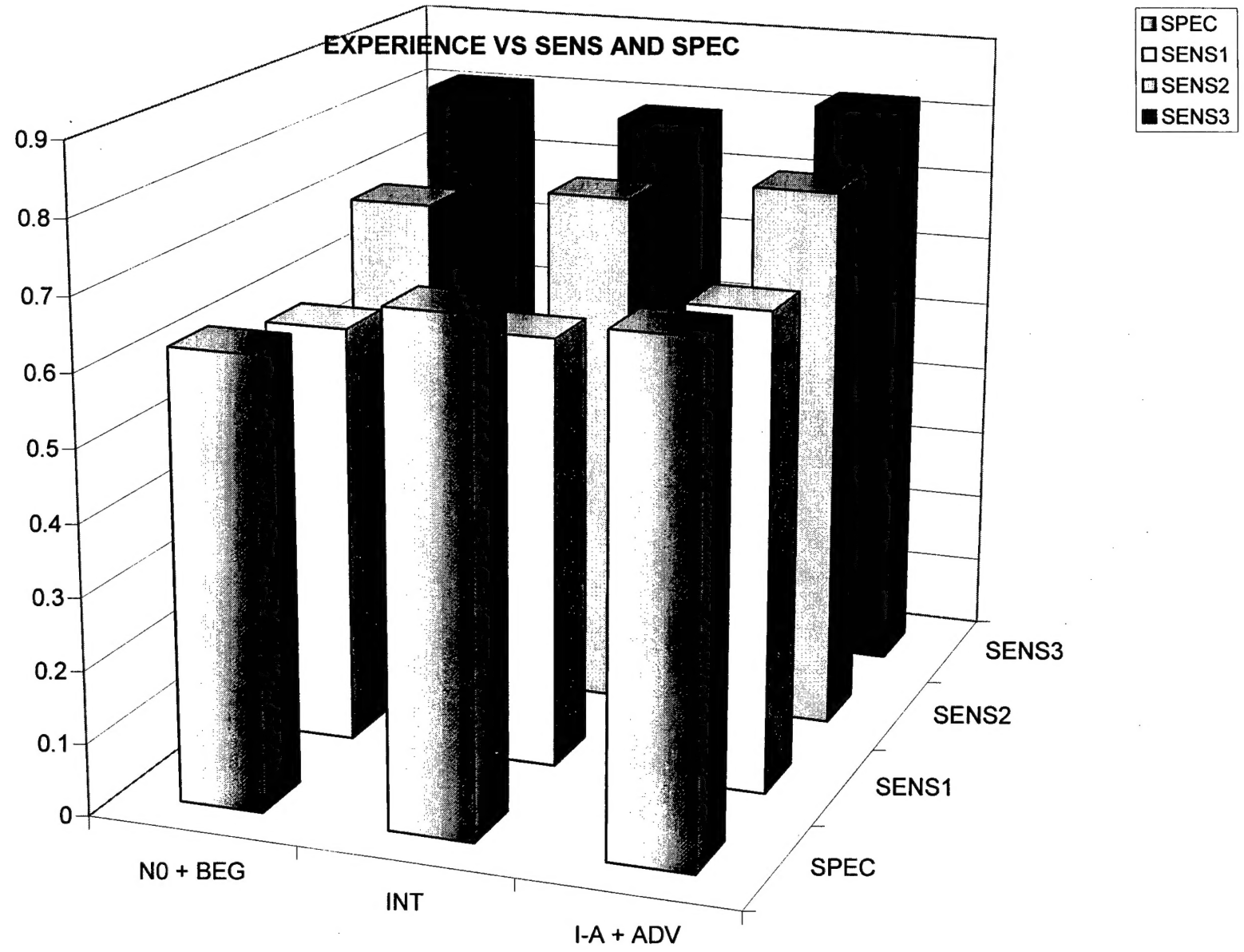
| | | | | | | |
|---|---|---|---|---|---|---|
| 39 | 54 | 17 | 7 | | 20 | 2 |
| 34.1 | 53 | 19 | | | 28 | |
| 43.2 | 48 | 2 | 3 | | 47 | |
| 26.2 | 47 | 1 | 1 | | 50 | 1 |
| 26.1 | 45 | 20 | 1 | | 33 | 1 |
| 91 | 45 | 7 | 1 | | 45 | 2 |
| 56 | 44 | 21 | 12 | | 23 | |
| 65 | 43 | 13 | 6 | | 36 | 2 |
| 50 | 40 | 12 | 6 | | 41 | 1 |
| 59 | 40 | 11 | 7 | | 42 | |
| 34.2 | 32 | | | | 68 | |
| 16 | 30 | 19 | 46 | | 5 | |
| 46 | 30 | 7 | 7 | | 54 | 2 |
| 47 | 20 | 14 | 61 | | 4 | 1 |
| 38 | 18 | 12 | 11 | | 54 | 5 |
| 11.1 | 15 | 2 | 1 | | 81 | 1 |



NEGATIVE CASES SORTED BY NUMBER OF OBSERVERS WHO ANSWERED CORRECTLY

| CASE# | 1.1 | 1.2 | 1.3 | 2 | 3 | 4 (blank) |
|---|---|---|---|---|---|---|
| 35 | | | | 100 | | |
| 66 | | | | 98 | 2 | |
| 28 | | | | 96 | 2 | 2 |
| 55 | | | | 95 | 5 | |
| 84 | | | | 93 | 7 | |
| 20 | | | | 92 | 8 | |
| 76 | | | | 92 | 5 | 3 |
| 61 | | | | 91 | 9 | |
| 68 | | | | 91 | 8 | 1 |
| 49 | | | | 90 | 9 | 1 |
| 97 | | | | 89 | 8 | 3 |
| 2 | | | | 88 | 12 | |
| 72 | | | | 88 | 12 | |
| 75 | | | | 88 | 11 | 1 |
| 32 | | | | 87 | 13 | |
| 51 | | | | 87 | 13 | |
| 54 | | | | 82 | 18 | |
| 48 | | | | 81 | 16 | 3 |


R MLO
L
#1
#2
#3
#4
#5
#6
#7
#8
#9
#10
#11
#12
R CC
lat↑L
low
Suspicion
high
① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩
NRML
ABNL
Lesion type
mass-S
mass-ID
mass-C
ASD
ARD
calc
other
For grading only


EXPERIENCE VS SENS AND SPEC
SPEC
SENS1
SENS2
SENS3
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
N0 + BEG
INT
I-A + ADV
SPEC
SENS1
SENS2
SENS3